WAISserver™ Release Notes

Version 2.0
October, 1994

A Wide Area Information Server System

For more information about WAIS products, contact **info@wais.com**
For WAIS customer support, contact **support@wais.com**

WAIS Inc.
1040 Noel Drive
Menlo Park, California 94025
(415) 617-0444

Printed in the United States of America.

Table of Contents

About This Release

Components and Bundled Products

The WAISserver software, Version 2.0, is a suite of programs designed to build, serve, monitor, and maintain a WAIS Database System. It consists of the `waisserver`, `waisparse`, `waisindex`, `waislookup`, `waisdelete`, and `waisreporter` programs.

With the release of Version 2.0, the WAISserver product is bundled with a suite of customization tools.

- A new product, WAISgate Version 2.0, enables the WAISserver administrator to connect a WAIS server with the World Wide Web.
- A new product, the Custom Parser Toolkit Version 2.0, provides instructions and code for writing custom parse formats.
- A new product, the Client Toolkit Version 1.0.1, provides instructions and code for writing new WAIS clients.
- A new product component, the `waisserver` filtering API, Version 1, lets the WAISserver administrator add specialized data retrieval and manipulation capabilities to the `waisserver` program.

Compatibility

Version 2.0 of the WAISserver software is fully compatible with all previous versions.

However, Version 2.0 of the WAISgate product is not compatible with previous, unbundled versions of WAISgate. In particular, `.html` files generated with an earlier version of WAISgate will not be properly read by Version 2.0 of WAISgate and must be regenerated.

Highlights

With the release of Version 2.0, the WAISserver software offers support for larger, more diverse, and more volatile data collections.

Special highlights include:

- **Support for Z39.50V2 protocol.**

  The WAISserver software now supports the WAIS Profile of Version 2 of the ANSI NISO Z39.50 protocol, also known as

Z38.50-1992 or Z39.50V2. The WAISserver software remains compatible with Version 1 of the Z39.50 protocol.

- **Better Administration Support.**

  WAISserver Version 2.0 includes a number of new and improved features designed to make WAISserver administration easier. In particular, incremental indexing re-indexes only the changes that have been made to a WAIS database since the last indexing—without re-indexing the whole database and without suspending service. Similarly, a variety of new custom configuration features are now available. For instance, custom configuration files make it possible to assign run-time parameters on a per-database basis.

  New flexibility has been added to WAISserver administration with the introduction of the new multi-database and external database capabilities. Now, multiple databases may be served as if they were one database and non-WAIS databases can be served by a WAIS server.

New Features and Capabilities

Incremental Indexing

If you need to add, delete, or modify documents in a WAIS database, Version 2.0 offers a variety of WAIS commands that allow you to update your database index without re-indexing all of the data and without requiring that users stop using the database. This process, known as incremental indexing, saves a great deal of time over the old method of re-indexing.

Several new features have been added to implement incremental indexing. The **`waisindex`** command now has an **`-append`** option, which is useful when new files are added to a collection of data or when old files are modified. Also, there is a new **`waisdelete`** command that selectively ignores all references to designated files in the WAIS index of a database.

For those who are administering a very volatile set of information, more fine-grained control over updates is offered. The **`-lock`**, **`-unlock`**, **`-replacewith`**, and **`-finalize`** options to **`waisindex`** are advanced features.

All of these incremental indexing features are detailed in Chapter 4, "Building a WAIS Database," and Chapter 8, "Command Reference," of the *WAISserver System Administration Manual*, Version 2.0.

In addition, the new **-merge** and **-tmp** options to **waisindex** allow you to perform the component operations of an **-append**. These are documented in Chapter 9, "Command Reference," of the *WAISserver System Administration Manual,* Version 2.0.

Multi-Databases

As of Version 2.0, the WAISserver software is capable of searching several normal databases as if they were one big database. A database configured in this manner is called a multi-database.

The benefits of a multi-database are numerous:

- **size and speed**

  A very large collection of data is often more easily managed if split up between two or more directories — possibly on separate disks. Splitting a database into two or more portions is also desirable because it cuts down on indexing time and space requirements when updating or modification is required.

- **flexibility**

  A multi-database appears to be a single database to WAIS client users. However, depending on how it is set up, a multi-database can offer users the greater flexibility of searching each component database separately or searching them all as a multi-database.

- **accuracy**

  The WAISserver multi-database mechanism ensures that document relevance scoring is merged across the component database searches.

Instructions for configuring a multi-database are given in Chapter 6 of the *WAISserver Administration Manual,* Version 2.0.

Serving External Databases

An external database is any non-WAIS database, such as an RDBMS, which has its own way of storing and retrieving data. An external database may reside on a remote machine, or it may be stored on a machine at your site.

Enhancements introduced with Version 2.0 enable the WAISserver software to serve a data collection even if that data is controlled by another database management system. Specifically:

- The **waisparse** program can now read a stream of input from standard input by using the + option. This makes it possible to parse and index data that does not reside in a UNIX file system.

- The **waisserver** program can now be easily configured to invoke custom filter programs, which retrieve and optionally modify external data. This makes it possible for the WAISserver to return external data in response to a retrieval request.
- A new configuration file scheme makes it easy to declare such filter programs.

Instructions for serving an external database may be found in Chapter 6 of the *WAISserver Administration Manual,* Version 2.0.

WAISgate

The World-Wide Web (WWW) is a popular Internet browsing and retrieval tool based on hypertext links between information sources. The new WAIS Inc. WAISgate product is a gateway between WAIS and WWW. If you are serving a WAIS database, WAISgate can greatly expand your user base by making your WAIS database(s) accessible from the WWW. Similarly, if you are serving a collection of Web pages, WAISgate brings database searching capabilities to Web client users.

Like WAIS, the World-Wide Web operates on a client-server model, but the two currently have different communication protocol and data format standards. WAISgate overcomes these differences. WAISgate receives WWW requests and forwards them to your WAIS server, effectively creating a gateway between a WWW server and a WAIS server. WAISgate can handle requests from any WWW clients that use Forms. This includes X-Mosaic, the most widely used WWW client, and LYNX, its character-based version.

Instructions for using WAISgate can be found in Chapter 6 of the *WAISserver Administration Manual,* Version 2.0 and in the file **/wais/current/waisgate/INSTALL**.

However, Version 2.0 of WAISgate is not compatible with previous, unbundled versions. In particular, **.html** files generated with an earlier version of WAISgate will not be properly read by Version 2.0 of WAISgate and must be regenerated. After installing the WAISserver software, follow the instructions in **/wais/current/waisgate/INSTALL** to install Version 2 of the **waisgate** and **waisexec** programs. Then use the **waisgate** program to regenerate Web **.html** files from your WAIS **.src** files.

Custom Parser Toolkit

A new product, the Custom Parser Toolkit, is bundled with WAISserver Version 2.0.

A large number of built-in parse formats are included in the WAISserver software. These parse formats can handle many standard file-based data formats. When a new file format is encountered, one of two strategies may be used to parse the new format. First, the data may be converted to one of the existing WAIS parse formats. Second, when conversion is not possible, a new parser can be developed by using the Custom Parser Toolkit. The Custom Parser Toolkit is a well-documented C program that includes source code examples and pseudo-code for all the customized functions you may need to write in order to create a new parse format. Alternatively, WAIS Inc. can provide custom programming services to develop and maintain specialized parsers to meet the needs of individual publishers on a contract basis.

The Custom Parser Toolkit is further described in Chapter 6 of the *WAISserver Administration Manual,* Version 2.0. Detailed instructions on how to use the Custom Parser Toolkit may be found on-line in the file **/srvr-2-0-*platform*/parser-toolkit/parser-manual.txt**, where *platform* represents the platform on which you are running the WAISserver software (e.g., sunos, ultrix).

Client Toolkit

The Client Toolkit included with this release is designed to help client writers get a fast start when developing new WAIS clients. It contains source code and on-line documentation of an Application Programmers Interface (API) for running both Version 1 and Version 2 of the Z39.50 information retrieval protocols. Also included in the toolkit is a demonstration program that shows the use of the client protocol toolkit API.

WAIS Inc. urges client writers to update their clients to the new protocol version and to develop new clients using Z39.50V2. This new protocol standard includes the ability to retrieve an unlimited number of documents, whereas the older version has strict limits based on the protocol buffer size — generally around 200 documents.

Documentation for the Client Toolkit can be found on-line in the directory **`/wais/current/client-toolkit/`**

License Control

With Version 2.0, a new license control mechanism is introduced. This mechanism requires that an encoded key be inserted into a global configuration file in order to turn on the WAISserver software. Thereafter, the software remains executable until the purchased software license term expires. Timely license renewal payments ensure uninterrupted software availability.

When you receive your copy of Version 2.0, please call the WAIS Inc. support line at (415) 617-0444 and ask for your software license key.

Enhancements and Modifications

Search Engine Enhancements

Version 2.0 adds several enhancements to the WAIS Search Engine.

Date and Numeric Ranges

WAISserver Version 2.0 extends the capabilities of fielded searches to handle searching for dates or positive integers that fall within user-defined ranges. These capabilities are fully documented in Chapter 6 of the *WAISserver System Administration Manual.*

As an example of a date range search, the query

```
date >= 10/14/94 AND name = Tiernan
```

searches for all documents in which the date field value is greater than or equal to October 14, 1994 and the name field value is "Tiernan". Similarly, the query

```
date = 10/1/94 TO 10/30/94 AND height >= 20 AND name =
Tiernan
```

specifies both ends of the date range (inclusive) and adds a requirement that the height field value fall within the range of positive integers greater than or equal to 20.

Literal Phrases

End users may now search for exact phrases simply by enclosing them in within quotation marks. This is known as a *literal.* For example, the query

```
"glass darkly"
```

returns only documents that contain this phrase. This is equivalent to the more cumbersome query syntax

```
glass ADJ darkly
```

where **ADJ** means "adjacent to".

Literal phrases may be combined with any other query syntax. For example, the query

```
"mirror dimly" OR "glass darkly"
```

combines the Boolean operator **OR** with two literals.

Expanded Boolean Syntax

The set of Boolean operators recognized by the WAIS parser has been expanded to use the C-like operators **&&** and **||**. These are equivalent to **AND** and **OR**, respectively. Thus, the queries

```
bonnie && clyde || truth || dare
```

and

```
bonnie AND clyde OR truth OR dare
```

are equivalent.

Thesaurus Not Included

A previously advertised feature known as a Thesaurus and intended to aid query term expansion is not yet available. To avoid confusion, the description of this feature that appeared in the document entitled *Technical Description, Release 1.1* is not included in the Version 2.0 *Technical Description.*

Enhancements to `waisserver`

Version 2.0 of the WAISserver software includes many enhancements to the **`waisserver`** program that support more varied and flexible database administration. An overview of these enhancements is given below.

Customized Help and Notices Files

The administrator may add **`index.hlp`** and **`index.not`** files to any database index directory. When these files are present, they are used to provide end users with on-line help and on-line notices.

Local and Global Configuration Files

A WAIS configuration file defines various run-time parameters used by a **`waisserver`** process when handling requests against a WAIS database. Both local, or database-specific, and global, or shared, configuration files are supported. If you have created a global configuration file, its parameter values are the defaults. For any particular database, they are overridden by the parameter values given in the local configuration file, if there is one.

Multi-Databases

The waisserver and waislookup programs are now capable of searching multiple databases as one. This mode of operation is simply configured by creating a separate index directory for the multi-database and putting in it a local configuration file that contains a multi-database declaration.

Filter Programs

While handling a retrieval or relevance feedback request from a client process, the **waisserver** program is now capable of calling an external filter program. These custom programs retrieve and optionally modify or otherwise act on data. This new feature may be activated either by using the new **waisserver -filter** switch or by placing one or more filter declarations in a configuration file.

Reading Arguments from a File

The new **waisserver -args** switch enables it to read its arguments from a file instead of from the command line. This is particularly useful when waisserver processes are invoked by the **inetd** daemon because some **inetd** implementations limit the length of a command in the inetd configuration file.

Timeout for Idle Clients

The new **waisserver -t** option specifies how long a server process will wait before closing its connection with an idle client process.

All of these new features are documented in Chapter 6 of the *WAISserver Administration* manual, Version 2.0, with the exception of the **-args** option, which is documented in Chapter 5. In addition, the optional **waisserver** command switches **-config**, **-filter**, **-args**, and **-t** are documented in Chapter 9, the Command Reference.

Enhancements to **waisparse**

The **waisparse** program contains several enhancements that support some of the new, broader capabilities of Version 2.0.

First, **waisparse** has new flexibility in reading its input. If called with the **-** switch, as in

```
% find /dev/src -name '*.[ch]' -print | waisparse -
```

waisparse reads from standard input a list of files whose contents it then parses. For example, the above command line creates a list of all the **.c** and **.h** files in the **/dev/src** directory and pipes that list to **waisparse**.

If **waisparse** is called with the **+** switch, as in

```
dumpRDBMSdata | waisparse +
```

it expects input as a stream from standard input. The **+** option is generally used when piping data from an external source into waisparse. For example, in the command line above,

dumpRDBMSdata is a program that dumps data out of an RDBMS system.

Second, **waisparse** is now able to parse formats for the hyper-text markup language (HTML), which is the data format used by the World-Wide Web. This **waisparse** enhancement supports the new WAISgate product component.

Finally, **waisparse** can now handle parse formats that define multiple fields per line. This enhancement is essential when handling some fielded data formats -- especially those from external databases.

Enhancements to waisindex

With the release of Version 2.0, an enhanced **waisindex** program makes WAIS database administration easier.

First, the new **-nice** option allows the administrator to set the process priority for waisindex. It takes one argument between -20 and 20, the higher the number, the "nicer" the process, in the sense of relinquishing more resource time to other processes. For more information, see the **waisindex** command reference in Chapter 9 of the *WAISserver Administration Manual*, Version 2.0.

Second, **waisindex** has been given a whole suite of new options to support incremental indexing. See the documentation for the **-append**, **-lock**, **-unlock**, **-tmp**, **-merge**, and **-replacewith** switches in the **waisindex** command reference, along with the **waisdelete** command reference in Chapter 9 of the *WAISserver Administration Manual*, Version 2.0. Also see the discussion of incremental indexing in Chapter 4, "Building a WAIS Database."

The waislookup Command

The **waislookup** command invokes a simple, interactive WAIS interface. It has been improved for this release and now replaces the freeware **waissearch** command.

The **waislookup** program can now be used either as a local or remote client process. As a local client, it bypasses **waisserver** and consults index files and retrieves data directly. As a remote client, it interacts with a **waisserver** process, just as any WAIS client does.

In addition, waislookup can now be used to search remote databases by using the following syntax

waislookup -d *database-name@server-host:host-port*

As with the rest of this release, **waislookup** supports both the 1992 and the 1988 versions of the Z39.50 protocol.

The **`waislookup`** command is fully documented in Chapter 8, "Clients and Queries" and in Chapter 9, "Command Reference" of the *WAISserver Administration Manual,* Version 2.0.

Optimizations

With the release of Version 2.0, the WAISserver software has been improved to provide more robust disk and memory overflow handling and to use a faster filename table searching algorithm.